IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-6, Issue-4 E-ISSN: 2347-2693

A Study on Crowd Detection and Density Analysis for Safety Control

Mayur D. Chaudhari[1*], Archana S. Ghotkar[2]

[1*]Computer, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, India
[2] Computer, Pune Institute of Computer Technology, Savitribai Phule Pune University, Pune, India

[*]*Corresponding Author: chaumayu@gmail.com, Tel.: +91 8275339994*

Available online at: www.ijcseonline.org

Received: 14/Mar/2018, Revised: 20/Mar/2018, Accepted: 05/Apr/2018, Published: 30/Apr/2018

***Abstract*—** Most of the studies based on tracking individuals, crowd counting, finding the region of motion and crowd detection. Crowd detection and density estimation from crowded images have a wide range of application such as crime detection, congestion, public safety, crowd abnormalities, visual surveillance and urban planning. The purpose of crowd density analysis is to calculate the concentration of the crowd in the videos of observers. Pattern recognition technique helps to estimate the crowd detection count and density by using face and detection. The job of detecting a face in the crowd is complicated due to its variability present in human faces including color, pose, expression, position, orientation, and illumination. The counting performance has been steadily improved because of Deep Convolutional Neural Network..

***Keywords*—**Pattern Recognition, Computer Vision, Crowd Density Estimation, Detection, CNN

I. INTRODUCTION

In recent years, the human population is growing in extreme rate hence the growth has indirectly increased the incidence of the crowd. The motive of assembly has major result in wide-range assets and crowded behavior. There is a lot of interest in much scientific research in public service, security, safety and computer vision for the analysis of mobility and behavior of the crowd[1]. The job of detecting a face in the crowd is complicated because of showing variance human faces including color, pose, expression, position, orientation, and illumination.

Due to a crowded crisis, there are large crowds of confusion, consequence in pushing, mass-panic, stampede or crowd crushes and causing control loss[8]. Some examples of crowd tragedies, crushes, stampede, like Heavy rains killed 22 people and injured hundreds of others in the afternoon between Mumbai, Parel and Elphinstone Road 2017, 27 pedestrians died due to a stampede on the banks of Godavari river 2015 in southern Indian state of Andhra Pradesh as shown in Fig. 1, 32 people died and 26 others injured after the Stampede held on the occasion of Diwali at Gandhi Maidan 2014.

To prevent these fatalities, automatically detection of critical and unusual situations in the dense crowd is necessary. As a result, definitely will help, to make emergency controls and appropriate decisions for security and safety[1]. Crowd detection is one of the most challenging tasks in visual surveillance systems. This system can be used for detection and count people, crowd level and also alarms as the presence of the dense crowd.

Figure 1. A Crowd Image Example - Bank of Godavari, Datia District
[source-https://blogs.timesofindia.indiatimes.com/random-harvest/stampedes-as-indian-as-taj-mahal/]

The purpose of a crowded counting is to count the number of people in the crowded places[2]. There are some applications of crowd detection, such as (1) ***Safety Control*** - Video surveillance cameras for safety purpose in places such as sports stadium, shopping malls and airports have validated monitoring of crowd for behaviour analysis, congestion analysis, and anomaly detection. (2) ***Disaster Management*** - Crowd gathering such as music concerts and political rallies face the risk of disasters such as stampede. Need to use for early overcrowding detection. (3) ***Public Areas*** - There are many of public locations where crowd level may be high such as malls, stations, terminals and some others which may be affected by human health. (4) ***Visual Surveillance*** - Public

places such as playground and huge arena are very crowded hence this type of system may fail to detect an individual in the crowd. Visual Surveillance system helps to reduce the failure percentage by anomaly detection and alarming.

Crowd counting systems consist of various approaches such as detection and regression-based estimation. Detection based approach involves to segment and recognize each individual crowd scenes followed by counting with some classifiers[9]. Recently, Regression based approach used Convolutional Neural Network (CNN). ***Challenges*** - There are many challenges in crowd analysis such as occlusions, high clutter, contrast variations, non-uniform distribution of people, non-uniform illumination, low resolution, intra-scene and inter-scene variations[2].

Zhan *et al.*[10] and Junior *et al.*[11] studied and reviewed existing practices for general crowd analysis. Li *et al.*[12] surveyed different methods for such crowd scene analysis such as crowd motion pattern learning, crowd behaviour, activity analysis and anomaly detection in crowds. Loy *et al.*[13] provided a detailed description and comparison of crowd count based on video images. While comparing with other approaches CNN-based approaches have managed to extremely reduce error rates.

Paper is organized as follows, Section 2 describes Approaches to Crowd Detection System approaches. Section 3 surveyed on CNN based methods. Section 4 discussed CNN layers. Finally, the conclusion is presented in Section 5**.**

II. APPROACHES TO CROWD DETECTION SYSTEM

Crowd Detection System consists of Input Data, Approaches, Features and Conclusion as shown in Fig. 2. There are various approaches have been taken to handle the problem which can be broadly divided into detection based approaches, regression based approaches, density based approaches.

A. Detection based approaches

Detection model tries to determine the number of people by identifying a single person and their places at the same time. Jones and Snow *et al.*[14] described as a Scanning window pedestrian detector using spatiotemporal information[15]. Haar-like filters, absolute difference Haar filter and shifted difference filter are three types of filters which used for capturing moving objects. Adaboost learning algorithm is using to trained eight different pedestrian detectors for eight motions[16]. Besides, this algorithm is utilized to exploit both movement and appearance data to make and try to arrange the moving individual. Leibe *et al.* [17] presented an algorithm for pedestrian detection in crowded scenes uses an algorithm to combine local and global features in a potential top-down segmentation. Their experiments indicated that they are dependent on the system and pedestrians can be severely localized, even after severe overlapping.

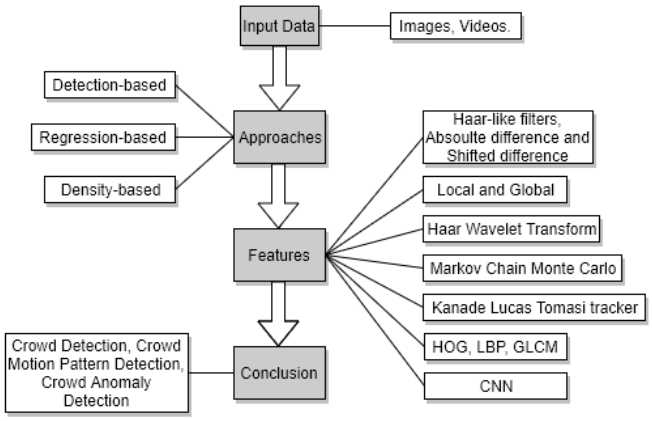

Figure 2. General Structure of Crowd Detection System

Lin *et al.* [18] proposed a detection technique for the crowded estimates through wavelet templates and vision-based technologies. The Haar Wavelet Transform (HWT) work was carried out to extract the feature's specification of head shape. Support Vector Machine (SVM) was carried out for a featured area so that it can be categorized as the presence and absence of head. This method was limited during complex situations when the head was not clear and proved heavy computational load on real-time applications [19]. Zhao and Nevatia [20] presented a 3D human shape model to recognize and look for crowded individuals. Their proposed technique depends on the head top detection by separating the foreground blobs. Likewise, a Maximum A Posteriori problem was formulated for identification and tracking problem of people. There is an occlusion problem to prevent the joint likelihood of different people based on the approach of Markov Chain Monte Carlo (MCMC).

Rabaud and Belongie [21] proposed a technique of dividing movements to be generated by numerous examples of a person in a crowd. Implemented a Kanade Lucas Tomasi (KLT) tracker[22] which is highly parallelized which used to extract an expensive arrangement of low-level features to identify the object which is in moving the state from the scene. In addition, KLT tracker is to distinguish the number of objects moving in a single scenario, the trajectory set is integrated with the temporal and spatial filter through the clustering method. Sidla *et al.* [23] presented a movement detection and tracking device to measure individual in very crowded situations. They developed an algorithm which recognizes human head-shoulder regions (Ω-like shape) and masked by region of interest (ROI) filter to recognize human in the crowd. The Kanade Lucas Tomasi (KLT) tracking point and Kalman filter are used to calculating the co-occurrence matrix feature vector for active size model to analyzed pedestrian movement. Detection based approach successful in the low-density crowd and affected in the high-

density crowd. Detection based approach successful in a low-density crowd and affected by a high-density crowd.

B. Regression based approaches

Regression based approach works on local image patches which extract mapping between features for counting purpose. There are various features to encode low-level information such as foreground features, edge features texture and gradient features. These methods capture local and global properties of the scene such as Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Gray Level Co-occurrence Matrices (GLCM) to improve results. After extracting local and global features, different regression techniques are applied such as linear regression, ridge regression, the neural network to learn to map for crowd counting purpose[2]. Idrees *et al.* [24] identified not a single feature and detection method is enough reliable to provide sufficient information to accurately calculate the presence of the high-density problem hence they proposed Fourier analysis along with head detection and SIFT interest point or some different methods to extract features.

C. Density based approaches

Density based approach tries to learn the linear mapping between local path features and corresponding object density maps. But, observing that it is difficult to learn linear mapping. Pham *et al.* [25] proposed to learn a non-linear mapping between local patch features and density maps. Random Forest Regression from multiple image patches is used to vote for densities of multiple target options.

III. CONVOLUTIONAL NEURAL NETWROK BASED METHODS

Convolutional Neural Network based methods consist of deep learning approaches for crowd detection and density analysis. CNN uses for learning non-linear functions from crowd images to counts. Various methods have been proposed in the literature as follows. CNN performs two types of methodology such as patch based which training based on patches of images of different sizes and whole image based which works on the whole image.

CNN was firstly applied by Wang *et al.* [7] and Fu *et al.* [26] were among the first who applied CNNs for the task of crowd density estimation. Wang *et al.*[7] applied end-to-end deep CNN regression model for counting people from high dense crowded images. He developed AlexNet network which replaced fully connected layer of 4096 neurons with single neuron for predicting crowd. His approach comes in patch-based inference process. Fu *et al.*[26] Classified the image into five classes: very high, high, medium, low and very low density instead estimating density maps. His approach comes in patch-based inference process. C. Zhang *et al.*[27] proposes to learn a map of pictures for crowding calculations and adapt this mapping to new target scenes for cross-scene counting. Initially, studied their network by training two objectives: The estimated objectives of predicting the density and density of the crowd. His approach comes in patch-based inference process.

Y. Zhang *et al.* [3] proposed Multi-Column Convolutional Neural Network (MCCNN) architecture allows the image to be arbitrary size or resolution. To model the density maps corresponding to heads of different scales it uses filters for each column of different sizes. His approach comes in whole image based inference process. Rather than the above techniques that use patch based inference process. Shang *et al.*[28] using CNN proposed an end-to-end count estimation technique. Rather than cropping the image into patches, their method uses the whole image as input and gives output the final crowd count. Zeng *et al.* [4] proposed a novel Multi-Scale Convolutional Neural Network (MSCNN) for single image crowd counting. It used extracting scale relevant features from crowd images using a single column network based on the multi-scale blob. Kang *et al.* [5] proposed CNN-pixel and FCNN-skip architecture. CNN-pixel is pixel-wise prediction using CNN. FCNNskip is Fully Convolutional Neural Network with skip branches. Produced the highest quality density map for localization tasks, with slight degradation for the counting task.

IV. OVERVIEW OF CNN LAYERS

Convolutional Neural Network consists of a sequence of layers. There are mainly 3 types of layers to build architecture such as the Convolutional layer, Pooling layer, and Fully-connected layer. Architecture consist of Input - Convolution - ReLU - Pooling - Fully Connected.

A. Input layer

The 32 x 32 x3 will hold the raw pixel estimations of the image of width 32, height 32, and 3 colour channels consist of Red, Green, Blue.

B. Covolutional layer

This will figure the output of neurons which are associated with the local area in the input, each computing a dot product of their weight and an input volume. This may bring about volume, for example, 32 * 32 * 12 for 12 filters. It convolves input frame with linear sliding filters to create response maps.

$$X_i = \sum_{i=0}^{i} w_i \, x_i + \, b$$

where, *Xi* is the feature map of input, w_i = [*Wi*1, *Wi*2,....., *Wik*] indicates filter, *xi* = input filter, *bi* indicates bias.

C. Rectified Linear Units layer

ReLU is the acronym of Rectified Linear Units. This layer applies the non-saturating activation function. The ReLU

function is $f(x) = max(0; x)$. This layer leaves the size of the volume unchanged 32 x 32 x 12.

D. Pooling layer

Pooling executes a down sampling operation along with width and height, resulting in volume such as 16 x 16 x 12. In next layer of pooling layer, it combines the outputs of neuron clusters at one layer into a single neuron. Max pooling uses the maximum value from each of a cluster of neurons at the prior layer. Average pooling uses the average value from each of a cluster of neurons at the prior layer.

E. Fully-Connected layer

The fully Connected layer will compute the class scores. High-level reasoning in the neural network is done by fully connected layers. Layer requires a fixed number of inputs and outputs to convert response maps as close to the ground truth[7].

V. DATASETS

INRIA dataset: INRIA dataset is available on http://pascal.inrialpes.fr/data/human/ The dataset is divided into two forms such as, the original image with the corresponding annotation file, and the positive images of 64x128 pixel format with negative images.

Mall dataset: Mall dataset is available on http://personal.ie.cuhk.edu.hk/∼ccloy/downloads_mall_dataset.html. It consists of over 60,000 individuals on road were labelled in 2000 video images.

Caltech pedestrian dataset: Caltech dataset is available on http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/. It consist of 64x128 pixel format of images shows human.

PETA dataset: PETA dataset is available on http://mmlab.ie.cuhk.edu.hk/projects/PETA.html. It consists of 19000 images, with the resolution of 17x39 to 169x365 pixels. Those 19000 images include 8705 humans, each annotated with 61 binary and 4 multi-class attributes.

VI. CONCLUSION

Crowd detection and density estimations are one of the challenging problems of computer vision and machine learning. There are three approaches such as Detection-based approach, Regression-based approach, Density-based approach. The deep learning model is very efficient for crowd counting and analysis where we discussed on some methods of Convolutional Neural Network which is our basic framework to learn efficient features for counting. It is an end-to-end training method which performs a whole image based inference. To get better performance of crowd counting, it requires large labelled dataset.

REFERENCES

[1] Sami Abdulla, Mohsen Saleh, Shahrel Azmin Suandi, Haidi Ibrahim, "Recent survey on crowd density estimation and counting for visual surveillance", Engineering Application of Artificial Intelligence 41 (2015) pp. 103-114, http://dx.doi.org/10.1016/j.e ngappai.2015.01.0 07

[2] Vishwanath A. Sindagi, Vishal M. Patel, "A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation", Pattern Recognition Letters, elsevier 2017.

[3] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghau Gao, Yi Ma, "Single-Image Crowd Counting via Multi-Column Convolutional Neural Network", In CVPR, IEEE, pp. 589-597.

[4] Lingke Zeng, Xiangmin Xu, Bolun Cai, Suo, Qiu, Tong Zhang, "Multi-Scale Convolutional Neural Networks for Crowd Counting", IEEE 2017.

[5] Di Kang, Zheng Ma, Antoni B. Chan, "Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks - Counting, Detection, and Tracking", IEEE 2017.

[6] Ankan Bansal, K S Venkatesh, "People Counting in high Density Crowds from Still Images", IEEE 2015.

[7] Chuan Wang, Hua Zhang, Liang Yang, Si Liu, Ziaochun Cao, "Deep People Counting in Extremely Dense Crowds", ACM 2015.

[8] Helbing, D., Brockmann, D., Chadefaux, T., Donnay, K., Blanke, U., Woolley-Meza, O., Moussaid, M., Johansson, A., Krause, J., Schutte, S., et al., 2014. "Saving human lives: what complexity science and information systems can contribute". J. Stat. Phys., 147.

[9] Zhao, T., Nevatia, R., Wu, B., 2008. "Segmentation and tracking of multiple humans in crowded environments". IEEE Trans. Pattern Anal. Mach. Intell. 30 (7), pp. 1198-1211.

[10] Zhan, B., Monekosso, D.N., Remagnino, P., Velastin, S.A., Xu, L.Q., 2008. "Crowd analysis: a survey. Machine Vision and Applications" 19, pp. 345-357.

[11] Junior, J.C.S.J., Musse, S.R., Jung, C.R., 2010. " Crowd analysis using computer vision techniques". IEEE Signal Processing Magazine 27, pp. 66-77.

[12] Li, T., Chang, H., Wang, M., Ni, B., Hong, R., Yan, S., 2015. "Crowded scene analysis: A survey. IEEE Transactions on Circuits and Systems for Video Technology" 25, pp. 367-386.

[13] Loy, C.C., Chen, K., Gong, S., Xiang, T., 2013. "Crowd counting and profiling: Methodology and evaluation, in: Modeling, Simulation and Visual Analysis of Crowds". Springer, pp. 347-382.

[14] Jones, M.J., Snow, D., 2008. "Pedestrian detection using boosted features over many frames". In: 19th International Conference on Pattern Recognition, 2008. ICPR 2008. IEEE, pp. 14, http://dx.doi.org/10.1109 /ICPR.2008.4761703.

[15] Viola, P., Jones, M.J., Snow, D., 2005. "Detecting pedestrians using patterns of motion and appearance". Int. J. Comput. Vis. 63 (2), pp. 153-161.

[16] Schapire, R.E., Singer, Y., 1999. "Improved boosting algorithms using confidencerated predictions". Mach. Learn. 37 (3), pp. 297-336

[17] Leibe, B., Seemann, E., Schiele, B., 2005. "Pedestrian detection in crowded scenes". In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1. IEEE, pp. 878-885, http://dx.doi.org/10.1109/CVPR.2005.272

[18] Lin, S.-F., Chen, J.-Y., Chao, H.-X., 2001. "Estimation of number of people in crowded scenes using perspective transformation". IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 31 (6), pp. 645-654

[19] Lin, S.-F., Lin, C.-D., 2006. "Estimation of the pedestrians on a crosswalk". In: International Joint Conference SICE-ICASE, 2006. IEEE, pp. 4931-4936, http:// dx.doi.org/10.1109/SICE.2006.314851

[20] Zhao, T., Nevatia, R., 2004. "Tracking multiple humans in complex situations". IEEE Trans. Pattern Anal. Mach. Intell. 26 (9), 1208-1221.

[21] Rabaud, V., Belongie, S., 2006. "Counting crowded moving objects". In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1. IEEE, pp. 705-711, http://dx.doi.o rg/10.1109/CVPR.2006.92

[22] Shi, J., Tomasi, C., 1994. "Good features to track". In: 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94. IEEE, pp. 593-600, http://dx.doi.org/1 0.1109/CVPR.1994.323794

[23] Sidla, O., Lypetskyy, Y., Brandle, N., Seer, S., 2006. "Pedestrian detection and tracking for counting applications in crowded situations". In: IEEE International Conference on Video and Signal Based Surveillance, 2006. AVSS'06. IEEE, p. 70, http://dx.doi.org/10.1109/AVSS.2006.91.

[24] Idrees, H., Saleemi, I., Seibert, C., Shah, M., 2013. "Multi-source multiscale counting in extremely dense crowd images", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2547-2554.

[25] Pham, V.Q., Kozakaya, T., Yamaguchi, O., Okada, R., 2015. "Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation", in: Proceedings of the IEEE International Conference on Computer Vision, pp. 3253-3261

[26] Fu, M., Xu, P., Li, X., Liu, Q., Ye, M., Zhu, C., 2015. "Fast crowd density estimation with convolutional neural networks". Engineering Applications of Artificial Intelligence 43, 81-88.

[27] Zhang, C., Li, H., Wang, X., Yang, X., 2015. "Cross-scene crowd counting via deep convolutional neural networks", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 833-841

[28] Shang, C., Ai, H., Bai, B., 2016. "End-to-end crowd counting via joint learning local and global count", in: Image Processing (ICIP), 2016 IEEE International Conference on, IEEE. pp. 1215-1219

[29] Ankan Bansal and K. S. Venkatesh, 2015. "People Counting in High Density Crowds from Still Images". International Journal of Computer and Electrical Engineering, pp. 316-324.

